Science Advances

Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation

--Manuscript Draft--

Manuscript Number:	ScienceAdvances-D-15-01678R1
Full Title:	Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation
Short Title:	Novel MS susceptibility loci
Abstract:	We conducted a genome-wide association study (GWAS) on multiple sclerosis (MS) susceptibility in German cohorts with 4,888 cases and 10,395 controls. In addition to associations within the MHC region, fifteen non-MHC loci reached genome-wide significance. Four of these loci are novel MS susceptibility loci. They map to the genes L3MBTL3, MAZ, ERG, and SHMT1. The lead variant at SHMT1 was replicated in an independent Sardinian cohort. Products of the genes L3MBTL3, MAZ, and ERG play important roles in immune cell regulation. SHMT1 encodes a serine hydroxymethyltransferase catalyzing the transfer of a carbon unit in the folate cycle. This reaction is required for regulation of methylation homeostasis, which is important for establishment and maintenance of epigenetic signatures. Our GWAS approach in a defined population with limited genetic substructure detected associations not found in larger, more heterogeneous cohorts, thus providing new clues regarding MS pathogenesis.
Article Type:	Research Article
Section/Category:	Life Sciences
Manuscript Classifications:	Life sciences; Neuroscience; Clinical neuroscience; Neurology; Neurological disorders; Multiple sclerosis; Genetics; Human genetics; Molecular genetics; Epigenetics
Keywords:	Multiple Sclerosis; Genome-wide association study; DNA methylation; L3MBTL3; MAZ; ERG; DLEU1; SHMT1
First Author:	Till Felix Malte Andlauer
Corresponding Author:	Bertram Müller-Myhsok Max-Planck-Institut fur Psychiatrie München, GERMANY
Corresponding Author E-Mail:	bmm@psych.mpg.de
Corresponding Author's Institution:	Max-Planck-Institut fur Psychiatrie
Corresponding Author's Secondary Institution:	
Order of Authors:	Till Felix Malte Andlauer
	Dorothea Buck
	Gisela Antony
	Antonios Bayas
	Lukas Bechmann
	Achim Berthele
	Andrew Chan
	Christiane Gasperi
	Ralf Gold
	Christiane Graetz
	Jürgen Haas
	Michael Hecker
	Carmen Infante-Duarte

	Matthias Knop
	Tania Kümpfel
	Volker Limmroth
	Ralf Linker
	Verena Loleit
	Felix Luessi
	Sven Meuth
	Mark Mühlau
	Sandra Nischwitz
	Friedemann Paul
	Matthias Pütz
	Tobias Ruck
	Anke Salmen
	Martin Stangel
	Jan-Patrick Stellmann
	Klarissa Stürner
	Björn Tackenberg
	Florian Then Berg
	Hayrettin Tumani
	Clemens Warnke
	Frank Weber
	Heinz Wiendl
	Brigitte Wildemann
	Uwe Zettl
	Ulf Ziemann
	Frauke Zipp
	Janine Arloth
	Peter Weber
	Milena Radivojkov-Blagojevic
	Markus Scheinhardt
	Theresa Dankowski
	Thomas Bettecken
	Peter Lichtner
	Darina Czamara
	Tania Carrillo Roa
	Elisabeth Binder
	Klaus Berger
	Lars Bertram
	Andre Franke
	Christian Gieger
	Stefan Herms

	Georg Homuth
	Marcus Ising
	Karl-Heinz Jöckel
	Tim Kacprowski
	Stefan Kloiber
	Matthias Laudes
	Wolfgang Lieb
	Christina Lill
	Susanne Lucae
	Thomas Meitinger
	Susanne Moebus
	Martina Müller-Nurasyid
	Markus Nöthen
	Astrid Petersmann
	Rajesh Rawal
	Ulf Schminke
	Konstantin Strauch
	Henry Völzke
	Melanie Waldenberger
	Jürgen Wellmann
	Eleonora Porcu
	Antonella Mulas
	Maristella Pitzalis
	Carlo Sidore
	Ilenia Zara
	Francesco Cucca
	Magdalena Zoledziewska
	Andreas Ziegler
	Bernhard Hemmer
	Bertram Müller-Myhsok
Order of Authors Secondary Information:	
Corresponding Author Secondary Information:	
Suggested Reviewers:	Stephen Sawcer, Prof. University of Cambridge sjs1016@cam.ac.uk
	Stephen Hauser, Prof. University of California San Francisco hausers@neurology.ucsf.edu
	Tomas Olsson, Prof. Karolinska Institutet Tomas.Olsson@ki.se

	Filippo Martinelli Boneschi, Dr. Ospedale San Raffaele martinelli.filippo@hsr.it
	Yurii Aulchenko, Dr. Novosibirsk State University yurii.aulchenko@gmail.com
Opposed Reviewers:	

Dear Dr Whetstine,

We would like to submit the revised version of our manuscript entitled “Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation”.

Thank you very much for the detailed and thoughtful reviews. We have answered and fulfilled all suggestions by the reviewers, and hope that the manuscript is now acceptable for publication in *Science Advances*.

With best regards,

Bernhard Hemmer and Bertram Müller-Myhsok

Names, telephone, and e-mail addresses for all authors:

T.F.M. Andlauer, +49-89-30622-222, till_andlauer@psych.mpg.de
D. Buck, +49-89-4140-7660, dorothea.buck@tum.de
G. Antony, +49 6426 8195940, gisela.antony@med.uni-marburg.de
A. Bayas, +49-821-400-3892, Antonios.Bayas@klinikum-augsburg.de
L. Bechmann, +49-391-67-17804, Lukas.bechmann@med.ovgu.de
A. Berthele, +49-89-4140-4673, achim.berthele@tum.de
A. Chan, +49-234-509-0, Andrew.Chan@rub.de
C. Gasperi, +49-89-4140-1, gasperi@lrz.tum.de
R. Gold, +49-234-509-2411, ralf.gold@rub.de
C. Graetz, +49-6131 17-3110, christiane.graetz@unimedizin-mainz.de
J. Haas, +49-6221/56-38980, Juergen.Haas@med.uni-heidelberg.de
M. Hecker, +49 381 494 5870, michael.hecker@rocketmail.com
C. Infante-Duarte, +49 30 450 539 055, Carmen.infante@charite.de
M. Knop, +49-89-30622-215, knop@psych.mpg.de
T. Kümpfel, +49 89 4400 74435, Tania.Kuempfel@med.uni-muenchen.de
V. Limmroth, +49 221 89073775, limmrothv@kliniken-koeln.de
R.A. Linker, +49 9131 8532187, ralf.linker@uk-erlangen.de
V. Loleit, +49-89-4140-1, verena.loleit@tum.de
F. Luessi, +49-6131 17-3110, felix.luessi@unimedizin-mainz.de
S.G. Meuth, +49 251 8346811, sven.meuth@ukmuenster.de
M. Mühlau, +49-89-4140-4661, mark.muehlau@tum.de
S. Nischwitz, +49-89-30622-403, slutz@psych.mpg.de
F. Paul, +49 30 450 539 705, Friedemann.Paul@charite.de
M. Pütz, +4964215865193, puetz@med.uni-marburg.de
T. Ruck, +49 251-83 46811, tobias.ruck@ukmuenster.de
A. Salmen, +49-234-509-0, anke.salmen@ruhr-uni-bochum.de
M. Stangel, +49-511-532 6676, Stangel.Martin@mh-hannover.de
J.P. Stellmann, +49-40-7410-54076, j.stellmann@uke.de
K.H. Stürner, +49-40-7410-22399, klarissa.stuerner@zmnh.uni-hamburg.de
B. Tackenberg, +49-6421-5865193, tackenbb@med.uni-marburg.de

F. Then Bergh, +49-341-9724307, Florian.ThenBergh@medizin.uni-leipzig.de
H. Tumani, +49-731-50063011, hayrettin.tumani@uni-ulm.de
C. Warnke, +49-211-81-08081, clemens.warnke@med.uni-duesseldorf.de
F. Weber, +49 6434 9190, f.weber@medicalpark.de
H. Wiendl, +49 251 83-4 68 10, heinz.wiendl@ukmuenster.de
B. Wildemann, +49 6221-56 7504, Brigitte.Wildemann@med.uni-heidelberg.de
U.K. Zettl, +49 381 494-9656, zettl@med.uni-rostock.de
U. Ziemann, +49 7071 29 82049, ulf.ziemann@uni-tuebingen.de
F. Zipp, +49 6131 17-7156, frauke.zipp@unimedizin-mainz.de
J. Arloth, +49-89-30622-354, arloth@psych.mpg.de
P. Weber, +49-89-30622-222, pweber@psych.mpg.de
M. Radivojkov-Blagojevic, +49 89 3187-2642, milena.radivojkov@helmholtz-muenchen.de
M.O. Scheinhardt, +49 451/500 – 2785, scheinhardt@imbs.uni-luebeck.de
T. Dankowski, +49 451/500 – 2782, dankowski@imbs.uni-luebeck.de
T. Bettecken, +49-89-30622-250, bettecken_05@psych.mpg.de
P. Lichtner, +49 89 3187-3530, lichtner@helmholtz-muenchen.de
D. Czamara, +49-89-30622-554, darina@psych.mpg.de
T. Carrillo-Roa, +49-89-30622-589, tania_carrillo@psych.mpg.de
E.B. Binder, +49-89-30622-586, binder@psych.mpg.de
K. Berger, +49-251-83-55650, bergerk@uni-muenster.de
L. Bertram, +49 451 879 29521, lars.bertram@uni-luebeck.de
A. Franke, +49 431 / 597 - 4138, a.franke@mucosa.de
C. Gieger, +49 89 3187-4106, christian.gieger@helmholtz-muenchen.de
S. Herms, +41 61 328 5019, stefan.herms@uni-bonn.de
G. Homuth, +49 3834 86 5873, georg.homuth@uni-greifswald.de
M. Ising, +49-89-30622-430, ising@psych.mpg.de
K.H. Jöckel, +49 201 / 92239 200, k-h.joeckel@uk-essen.de
T. Kacprowski, +49-3834-86-5874, tim.kacprowski@uni-greifswald.de
S. Kloiber, +49-89-30622-439, stkloiber@psych.mpg.de
M. Laudes, +49-431-597-1380, Matthias.Laudes@uksh.de
W. Lieb, +49 431 597-3677, Wolfgang.Lieb@epi.uni-kiel.de
C.M. Lill, +49 451 879 295 22, christina.lill@uni-luebeck.de
S. Lucae, +49-89-30622-1, lucae@psych.mpg.de
T. Meitinger, +49 89 3187-3294, Meitinger@helmholtz-muenchen.de
S. Moebus, +49 201 / 92239 230, susanne.moebus@uk-essen.de
M. Müller-Nurasyid, +49 89 3187 4104, martina.mueller@helmholtz-muenchen.de
M.M. Nöthen, +49 228-6885-400, markus.noethen@uni-bonn.de
A. Petersmann, +49 3834 86 56 70, astrid.petersmann@uni-greifswald.de
R. Rawal, +49 89 3187-1195, rajesh.rawal@helmholtz-muenchen.de
U. Schminke, +49 3834 86-6819, ulf.schminke@uni-greifswald.de
K. Strauch, +49 89 3187-2838, strauch@helmholtz-muenchen.de
H. Völzke, + 49 3834 86 7541, voelzke@uni-greifswald.de
M. Waldenberger, +49 89 3187-1270, waldenberger@helmholtz-muenchen.de
J. Wellmann, +49 251 / 83 55648, wellmann@uni-muenster.de
E. Porcu, +39 070 675465, eleonoraporcu@gmail.com
A. Mulas, +39 070 675465, antonellamulas6@gmail.com

M. Pitzalis, +39 070 675465, maristella_77@yahoo.it
C. Sidore, +39 070 675465, scarlino81@gmail.com
I. Zara, +39 070 675465, zara@crs4.it
F. Cucca, +39 070 675465, fcucca@uniss.it
M. Zoledziewska, +39 070 675465, madzia.zoledziewska@gmail.com
A. Ziegler, +49 (451) 500-2780, ziegler@imbs.uni-luebeck.de
Corresponding authors:
B. Hemmer, +49-89-4140-4601, hemmer@tum.de
B. Müller-Myhsok, +49-89-30622-246, bmm@psych.mpg.de

Response to reviewers

We thank the reviewers for their positive and helpful comments. We have incorporated all suggested changes and address the reviewers' comments in detail below.

Reviewer 1 Comments to Author:

1. *As regards statistical power, your sentence ("We expected to have sufficient power to detect novel associations with moderate effect sizes in our data set of roughly 4000 cases showing low population stratification") should be better framed with a more detailed range of effect size identified with used sample size.*

We agree with the reviewer that this statement needed clarification. We have added power analyses for two different effect sizes covering the range relevant for this manuscript (page 7, last sentence of the introduction).

2. *I would smooth in the discussion the conclusions of the predominant role of epigenetic regulatory mechanisms in MS, or add additional functional work to reinforce this strong statement. Apart from SHMT1, the functional link of identified variants with epigenetic mechanisms of the other 4 variants should be demonstrated and it is only based on the function of the genes pointed by the variants.*

We agree with the reviewer that we cannot directly prove a link between MS susceptibility and the epigenetic mechanisms, which the genes we describe in our study are connected to. Nonetheless, we believe that it is worth describing the connections of the novel genes to epigenetic mechanisms, especially in the case of *SHMT1*. In our opinion it is beyond the scope of this study to demonstrate which of these functions are responsible for MS susceptibility. The mechanistic link between these genes and MS should be addressed in future work. We have therefore smoothed out the discussion, shortened the relevant paragraph and stated more clearly that it constitutes a hypothetical discussion (page 15, second paragraph (*While a clearer picture has…*), until the end of the discussion).

Reviewer 2 Comments to Author:

1a. *The authors suggestion that their observed association with rs2812197 is novel is not sustainable. These same authors have recently published genomewide significant association with rs806349 and thereby have already established this association (authors ref 28). In this new study the authors show convincingly that association at rs806349 is secondary to rs2812197 but this does not make the association with rs2812197 a novel finding. The authors should re-word their paper to indicate that they have found 15 genomewide significant associations 11 of which are known and 4 of which are novel OR should indicate that in preparation of this work one of the 5 novel findings has been independently confirmed (by them!).*

We apologize that our wording regarding rs2812197 was misleading. We were trying to point out that a variant in *DLEU1* reached genome-wide significance for the first time in a GWAS. However, we understand the reviewer's concern that this does not qualify as a novel association and therefore have changed the wording throughout the text.
We now state within the results section: "Variants at fifteen loci outside the MHC region showed genome-wide significance. Ten of these loci have already been established in previous large MS GWAS. One more locus, *DLEU1*, was only recently confirmed to be associated with MS in a candidate SNP study. The remaining four signals are thus novel candidates for MS susceptibility loci." (page 9, second paragraph).
Accordingly, we now write in the discussion: "Four of the 15 non-MHC loci have not been found to be associated with MS in previous studies. One more locus, *DLEU1*, did not reach genome-wide significance in previous GWAS but has recently been confirmed as MS-associated in a candidate SNP study." (page 13, last paragraph).
We conclude from the reviewer's comment that it is still worth examining *DLEU1* in detail, as we can show that published variants for the locus are most likely not the causal ones. We have therefore moved our results regarding the *DLEU1* locus to the end of the results section (page 12, last paragraph). Furthermore, we have added a detailed comparison to previous

findings in the new table 5, which summarizes our analyses regarding *DLEU1*. We have removed this comparison from the discussion to avoid repetitions.

1b. *Furthermore in the WTCCC2 MS GWAS (the authors ref 6) a total of 102 associated SNPs were identified in the screening phase one of which was rs806321 from the DLEU1 region. This SNP is in strong LD with rs2812197, the authors should thus also mention ref 6 as previously having implicated this region/gene.*

We have now also incorporated this variant in the results (page 12, last paragraph), table 5, and fig. S6.

1c. *In the discussion the authors say "Five of the primary signals reaching genome-wide significance in the pooled analysis of DE1 and DE2 have not been established in previous GWAS as MS susceptibility loci so far or have been attributed to other variants, as is the case for DLEU1" Given that rs806321 and rs2812197 have an R squared of 0.65 it is not reasonable for the authors to suggest that previous studies implicating DLEU1 have identified independent variants. This suggestion should be removed.*

We have removed this sentence from the manuscript. We now write: “Four of the 15 non-MHC loci have not been found to be associated with MS in previous studies. One more locus, *DLEU1*, did not reach genome-wide significance in previous GWAS but has recently been confirmed as MS-associated in a candidate SNP study.” (page 13, last paragraph).

1d. *Again I am not sure I follow what is meant by the statement "The DLEU1 locus contained a second signal, rs9591325," do the authors mean that after conditional analysis on the lead SNP this SNP shows association? Or do they mean that this is another SNP in the region which shows association but is only modestly correlated with the lead SNP? This needs to be clarified. The authors note that association with rs9591325 is reduced after conditioning on rs2812197, but should also describe the reverse, is rs2812197 still associated after rs9591325 and by how much? The R-squared between these SNPs is modest. Genetically it would seem that rs2812197 is a stronger candidate than rs9591325 but the authors observe that rs9591325 lies in a more functionally active region. However I think their statement that rs9591325 is therefore the most likely to be causal is not fully justified. It would seem reasonable to explain that their study lacks the power to fine map this region, which might of course contain more than one associated variant, and that rs9591325 has more functional data to support its relevance, but to claim it is the most likely causal variant is unjustified.*

We apologize that the description of rs9591325 was confusing. We changed the sentence as follows: “The *DLEU1* locus contains evidence for a second signal, rs9591325, [...]” (page 13, second paragraph). As the reviewer states, we did not have enough power in our study to prove that both signals are independent. We have added more detailed analyses to support our hypothesis that “The two signals were partially independent of each other” (page 13, second paragraph). As the reviewer suggested, we have conducted a detailed conditional analysis in both directions that is described in table 5. Finally, we now point out that we cannot fully separate between both signals in our study: “While rs2812197 shows the overall strongest association at *DLEU1*, the functional data indicate that rs9591325 might be either the actual or a second causal variant. Additional studies with larger sample sizes are required to fully answer this question.” (page 13, end of results section).

2. *Supplementary table S5 is very detailed but contains a lot of primary information. I think that at least the main results from this table should be included in the main text. The results for the known associated variants in the Sardinian population should be included. Details about the number of SNPs in each region and how many were genotyped can stay in the supplementary file. The results section "Additional novel candidate loci associated with MS" is largely a description of details in table S5, once this is in the main text this section can be radically reduced/rationalised.*

In order to address the reviewer’s suggestion, we have reduced the information included in fig. 1 and instead generated a new table (table 3) that contains part of the information from

former fig. 1 as well as the key data from former table S5. We have tried to balance the information included in the new table 3 to make it as informative as possible to the reader yet also make sure that it still fits onto an A4 page. The remaining information of former table S5 can now be found under the new name table S4. Furthermore, we have shortened the details described in the results section, as the reviewer suggested.

3. *Given that the authors found 58 variants with lower and 35 with higher OR, it is clear that they did in fact find more OR values were reduced than were increased. This difference may not have reached statistical difference from chance but 58 is certainly larger than 35. In which case I don't follow the statement "Thus we observed neither an increased number of lower nor of higher ORs than expected by chance (binomial test p-values 0.14 (CI 0.47-1.00) and 1.00 (CI 0.26-1.00), respectively)." Regression to the mean would predict that more OR should be reduced than would be increased and I think it would be more helpful to the reader to point out that the observed changes are in line with that expected after regression to the mean.*

We agree with the reviewer that our wording was convoluted here. Accordingly, we have revised the sentence: “58 of the variants had lower and 35 higher ORs in our data than in the published data set. It was expected to observe more signals with lower ORs than previously reported due to regression towards the mean.” (page 10, end of first paragraph).

4. *In describing the rs4925166 associated region the authors state that "The most strongly associated genotyped SNP, rs12946752, reached a p-value of 3.57×10-9." I don't understand this statement. Firstly rs4925166 is more strongly associated and secondly rs12946752 is a perfect proxy for rs4925166. This sentence should be clarified or removed, does this concern which SNPs were genotyped and which imputed? If so it is not clear which is which.*

In this sentence of our initial manuscript, we described the strongest association of a genotyped (and not imputed) SNP. We have followed the reviewer’s suggestion to condense the detailed description of results. Comparisons of results regarding imputed *vs.* genotyped variants have thus been removed from the results section. Instead, the number of genome-wide significant imputed and genotyped SNPs is listed in table S4.

5. *Have the authors undertaken systematic conditional analysis on the lead SNP in each of their genomewide significant regions? Do any of these contain genomewide significant secondary signals? If this has not been done it should be done and should be reported.*

We have done this analysis and added its results to the manuscript: “When conditioning for the lead variants at these four newly identified MS-associated loci, no evidence for secondary signals was found. Thus, the lead variants also constitute the most likely causal variants.” (page 12, second paragraph).

6. *The authors make no more than passing mention of the MHC. They should include a more detailed analysis of this region. Is the most associated SNP tagging 15:01? Have the authors imputed the classical HLA genotypes and tested these for association. Do they see the same HLA associations reported by the IMSGC (of which they seem to be a part). The authors should provide an MHC results section.*

We had not focused on a description of the MHC region simply because our findings were in line with studies by the IMSGC. However, as a result of the reviewer’s suggestion, we have conducted a thorough analysis of associations with *HLA* alleles. We have incorporated this analysis in a new results section titled “Associations within the MHC region” (beginning on page 8). We describe imputation of classical *HLA* alleles and their analysis in the text and the new table 2. The most strongly associated SNP is indeed tagging *DRB1*15:01*, as outlined in the revised manuscript.

7. *Have the authors looked for any effect on age at onset or evidence of interaction between loci, especially between MHC and non-MHC loci. It would seem logical to describe the results of such analysis.*

We have examined the age at onset but no signal was genome-wide significant. In addition, the top signal confirmed previous findings by the IMSGC. We have included this analysis in the first paragraph on page 9 as well as in the new fig. S3.

We analyzed potential interactions between loci, but no significant interactions were found. We summarize this result with the sentence: “We could not detect any significant interactions among the 15 top non-MHC variants or between them and SNP rs3104373 within the MHC region.” (page 9, end of second paragraph).

General remarks regarding the revised manuscript:
In addition to incorporating the modifications suggested by the reviewers, we have slightly shortened the text throughout the manuscript to improve readability.
Some questions by the reviewers required additional calculations in the statistical software R. In order to make results consistent and comparable throughout the manuscript, we have recalculated all associations of lead variants in R. Due to differences regarding how numbers are rounded by the two applications, some *p*-values have changed very slightly compared to previous calculations using PLINK. MAFs have now been consistently calculated on controls.
Previous tables S3 and S4 have been combined into one table S3. Accordingly, the designations of the additional supplementary tables have been adapted.
We came to the conclusion that the previous fig. S8F is informative enough for the reader to justify moving it to the main figures as fig. 3C. If you do not approve of this change, we could move this figure back to the supplementary material.

We believe that the manuscript has significantly improved with the help of the reviewers and hope that it is now acceptable for publication in *Science Advances*.

H1 FRONT MATTER

H2: Title

Long title:

Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation

Short title:

Novel MS susceptibility loci

Teaser:

Genome-wide study in Germans confirms known and identifies four novel multiple sclerosis gene loci.

H2:Authors

T.F.M. Andlauer[1,2]†, D. Buck[3]†, G. Antony[4], A. Bayas[5], L. Bechmann[6,7], A. Berthele[3], A. Chan[8], C. Gasperi[3], R. Gold[8], C. Graetz[9], J. Haas[10], M. Hecker[11], C. Infante-Duarte[12], M. Knop[1], T. Kümpfel[13], V. Limmroth[14], R.A. Linker[15], V. Loleit[3], F. Luessi[9], S.G. Meuth[16], M. Mühlau[2,3], S. Nischwitz[1], F. Paul[12], M. Pütz[17], T. Ruck[16], A. Salmen[8], M. Stangel[18], J.P. Stellmann[19], K.H. Stürner[19], B. Tackenberg[17], F. Then Bergh[20], H. Tumani[21,22], C. Warnke[23], F. Weber[1,24], H. Wiendl[16], B. Wildemann[10], U.K. Zettl[11], U. Ziemann[25], F. Zipp[9], J. Arloth[1,26], P. Weber[1], M. Radivojkov-Blagojevic[27], M.O. Scheinhardt[28], T. Dankowski[28], T. Bettecken[1], P. Lichtner[27,29], D. Czamara[1], T. Carrillo-Roa[1], E.B. Binder[1,30], K. Berger[31], L. Bertram[32,33], A. Franke[34], C. Gieger[35,36], S. Herms[37,38], G. Homuth[39], M. Ising[1], K.-H. Jöckel[40], T. Kacprowski[39], S. Kloiber[1], M. Laudes[41], W. Lieb[42], C.M. Lill[32,9], S. Lucae[1], T. Meitinger[27,29], S. Moebus[40], M. Müller-Nurasyid[43,44,45], M.M. Nöthen[37], A. Petersmann[46], R. Rawal[35,36], U. Schminke[47], K. Strauch[43,48], H. Völzke[49], M. Waldenberger[35,36], J. Wellmann[31], E. Porcu[50], A. Mulas[50,51], M. Pitzalis[50], C. Sidore[50], I. Zara[52], F. Cucca[50,51], M. Zoledziewska[50,51], A. Ziegler[28,53,54], B. Hemmer[2,3]*, B. Müller-Myhsok[1,2,55]*

H2:Affiliations

[1]Max Planck Institute of Psychiatry, Munich, Germany.

[2]Munich Cluster for Systems Neurology (SyNergy), Germany.

[3]Department of Neurology, Klinikum Rechts der Isar, Technische Universität München, Germany.

[4]Central Information Office KKNMS, Philipps University Marburg, Germany.

[5]Department of Neurology, Klinikum Augsburg, Germany.

[6]Department of Neurology, University of Leipzig, Germany.

[7]Institute of Medical Microbiology, Otto-von-Guericke University, Magdeburg, Germany.

[8]Department of Neurology, St. Josef Hospital, Ruhr-University Bochum, Germany.

[9]Department of Neurology, Focus Program Translational Neurosciences (FTN) and Research Center for Immunotherapy (FZI), Rhine-Main Neuroscience Network (rmn2), University Medical Center of the Johannes Gutenberg University Mainz, Germany.

[10]Department of Neurology, University Hospital Heidelberg, Germany.

[11]Department of Neurology, University of Rostock, Germany.

[12]NeuroCure Clinical Research Center, Department of Neurology, and Experimental and Clinical Research Center, Max Delbrück Center for Molecular Medicine, and Charité University Medicine Berlin, Berlin, Germany.

[13]Institute of Clinical Neuroimmunology, Ludwigs-Maximilians-Universität, Munich, Germany.

[14]Department of Neurology, Hospital Köln-Merheim, Germany.

[15]Department of Neurology, University Hospital Erlangen, Germany.

[16]Department of Neurology, Klinik für Allgemeine Neurologie, University of Münster, Germany.

[17]Clinical Neuroimmunology Group, Department of Neurology, Philipps-University of Marburg, Germany.

[18]Department of Neurology, Hannover Medical School, Germany.

[19]Institute of Neuroimmunology and MS and Department of Neurology, University Medical Centre Hamburg-Eppendorf, Germany.

[20]Department of Neurology and Translational Center for Regenerative Medicine, University of Leipzig, Germany.

[21]Department of Neurology, University of Ulm, Germany.

[22]Neurological Clinic Dietenbronn, Schwendi, Germany.

[23]Department of Neurology, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany.

[24]Neurological Clinic, Medical Park, Bad Camberg, Germany.

[25]Department of Neurology & Stroke and Hertie-Institute for Clinical Brain Research, Eberhard-Karls-Universität Tübingen, Germany.

[26]Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany.

[27]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany.

[28]Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Germany.

[29]Institute of Human Genetics, Technische Universität München, Germany.

[30]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia.

[31]Institut für Epidemiologie und Sozialmedizin der Universität Münster, Germany.

[32]Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), Institute of Neurogenetics and Institute of Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany.

[33]School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom.

[34]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Germany.

[35]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany.

[36]Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany.

[37]Institute of Human Genetics, University of Bonn, Germany.

[38]Department of Biomedicine, Division of Medical Genetics, University of Basel, Switzerland.

[39]Interfaculty Institute for Genetics and Functional Genomics, Ernst Moritz Arndt University and University Medicine Greifswald, Germany.

[40]Institute of Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University Duisburg-Essen, Germany.

[41]Department I of Internal Medicine, Christian-Albrechts University, Kiel, Germany.

[42]Institute of Epidemiology and Biobank popgen, Christian-Albrechts-Universität Kiel, Germany.

[43]Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany.

[44]Department of Medicine I, Ludwig-Maximilians-Universität, Munich, Germany.

[45]DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany.

[46]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Germany.

[47]Department of Neurology, University Medicine Greifswald, Germany.

[48]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany.

[49]Institute for Community Medicine, University Medicine Greifswald, Germany.

[50]Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy.

[51]Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Italy.

[52]Center for Advanced Studies, Research and Development in Sardinia (CRS4), Pula, Italy.

[53]Zentrum für Klinische Studien, Universität zu Lübeck, Germany.

[54]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa.

[55]Institute of Translational Medicine, University of Liverpool, United Kingdom.

†: These authors contributed equally to this work

*: To whom correspondence should be addressed: hemmer@tum.de and bmm@psych.mpg.de

H2:Abstract

We conducted a genome-wide association study (GWAS) on multiple sclerosis (MS) susceptibility in German cohorts with 4,888 cases and 10,395 controls. In addition to associations within the MHC region, fifteen non-MHC loci reached genome-wide significance. Four of these loci are novel MS susceptibility loci. They map to the genes *L3MBTL3*, *MAZ*, *ERG*, and *SHMT1*. The lead variant at *SHMT1* was replicated in an independent Sardinian cohort. Products of the genes *L3MBTL3*, *MAZ*, and *ERG* play important roles in immune cell regulation. *SHMT1* encodes a serine hydroxymethyltransferase catalyzing the transfer of a carbon unit in the folate cycle. This reaction is required for regulation of methylation homeostasis, which is important for establishment and maintenance of epigenetic signatures. Our GWAS approach in a defined population with limited genetic substructure detected associations not found in larger, more heterogeneous cohorts, thus providing new clues regarding MS pathogenesis.

H1 MAIN TEXT

H2:Introduction

Multiple sclerosis (MS) is an autoimmune disease of the central nervous system. Human leukocyte antigen (*HLA*) alleles, located within the major histocompatibility (MHC) region, have been identified as major genetic determinants for the disease (*1*, *2*). In addition, more than 100 non-MHC MS susceptibility variants have been described (*3*, *4*). Many of the genes carrying known susceptibility variants are involved in the regulation of either immune cell differentiation or signaling (*4*-*8*). However, as the heritability of MS is limited (*9*), environmental contributions to disease etiology are also important (*10*). Environmental influences can alter gene expression via epigenetic mechanisms (*11*). Epigenetic alterations, such as DNA methylation or histone

modifications, have been observed in tissue and cells of MS patients (*8, 12-14*). Nevertheless, the impact of epigenetic regulation in MS is not yet understood.

The known genetic variants outside the MHC region have predominantly been established in large international collaborative studies. In order to achieve large sample sizes with the power to detect associations, these studies have combined sample sets from diverse ethnic populations (*4-6*). So far, the variants affecting MS susceptibility identified in these studies account for only 25 % of disease heritability under an additive model of heritability (*3*), warranting for additional studies to fully unravel the genetic contribution to disease susceptibility. In contrast to the previously investigated large international cohorts, we have strived to examine the genetic contribution to MS susceptibility in a more homogenous population, focusing entirely on German cases and controls. The genetic substructure among Germans is low (*15*). We therefore expected to have sufficient power to detect novel associations with moderate effect sizes in a data set showing little population stratification. Indeed, with a total of 4,888 cases and 10,395 controls, we had 80 % power to detect genome-wide significant associations with an odds ratio (OR) of 1.2 involving common variants with a minor allele frequency (MAF) of 21 %. For rare SNPs (MAF 1 %) the power surpassed 80 % for an OR of 1.9.

H2: Results

Genome-wide association analyses

We recruited patients with either MS or clinically isolated syndrome (CIS) from MS centers throughout Germany and combined them with controls from several German population-based cohorts (table 1). After quality control (QC), this data set DE1 consisted of 3,934 cases and 8,455 controls (control/case ratio 2.15, table S1). We also compiled a second data set, called DE2, based on an independent group of German cases previously used in the IMSGC/WTCCC2 MS study (*5*) (table 1), and additional German controls, mostly from population-based cohorts. This data set DE2 contained 954 cases and 1,940 controls after QC (control/case ratio 2.03, table S2).

We observed only moderate population substructure within these data sets (figs. S1 and S2), confirming previous genetic analyses of the German population (*15*).

Both data sets were imputed separately to the 1000 genomes Phase I reference panel using SHAPEIT2 and IMPUTE2 (*16*-*18*). The resulting data sets contained over eight million high-quality variants with MAFs of at least 1% each. We conducted genome-wide association analyses (GWAS) on both data sets separately, using sex and the first eight multi-dimensional scaling (MDS) components of the genetic similarity matrix (GSM) as covariates, to control for any remaining population substructure. After assuring that the median genomic inflation of the two GWAS was in the expected range (table S3), results were combined using a fixed-effects pooled analysis. In this pooled analysis, the genomic inflation $\lambda_{1000,1000}$ outside the extended MHC region was 1.017 (table S3) (*19*).

Associations within the MHC region

The variant showing the strongest association in the pooled analysis of DE1 and DE2, rs3104373 (OR 2.90, confidence interval (CI) 2.72-3.09, *p*-value 1.3×10^{-234}), lies within the MHC region between the genes *HLA-DRB1* and *HLA-DQA1*. This single nucleotide polymorphism (SNP) is in strong linkage disequilibrium (LD) with the *HLA* allele *DRB1*15:01* (r^2 = 0.99) and thus corresponds to the established major MS risk locus (*1*, *2*). In order to confirm this finding, we imputed classical *HLA* alleles from our genotyping data (*20*). After QC, we obtained high-quality imputed alleles for a total of 3,966 cases and 8,329 controls from DE1 and DE2 (median accuracy 96.1 %, median call rate 97.4 %). Using step-wise conditional logistic regression (*1*, *2*, *5*), seven *HLA* alleles (table 2) reached genome-wide significance (*i.e.*, *p*-values < 5×10^{-8}). As expected, the most significantly associated allele was *DRB1*15:01* (OR 2.85 (2.66-3.06), *p*-value 1.0×10^{-191}). All seven alleles have been described as associated with MS in a recent detailed analysis of the MHC region (*2*).

Previous analyses of the MHC region have also identified associations between *HLA* alleles and the age at onset of the disease, mainly with *DRB1*15:01* (2, 5). We confirmed this finding in a subset of patients from our data set DE1. Age at onset was known for 1,519 patients, for 1,196 of them imputed *HLA* alleles were available. As the age at onset was not normally distributed, rank-based inverse normal transformation was applied. The *HLA* allele most strongly associated with transformed age at onset was *DRB1*15:01* (effect size -0.21, *p*-value 7.6×10^{-6}). When conducting a genome-wide analysis of transformed age at onset in all 1,519 patients, no variant passed the threshold for genome-wide significance (fig. S3A-B). The most strongly associated SNP was rs4959027 (effect size -0.20, *p*-value 1.5×10^{-7}; fig. S3C-D), which is in LD with *DRB1*15:01* ($r^2 = 0.72$). After conditioning for *DRB1*15:01* in the subset of cases with both age at onset and imputed *HLA* alleles available, the *p*-value of rs4959027 increased from 1.1×10^{-6} to 0.048. We conclude that our findings for the MHC region are very well in line with previous studies and concentrated further analyses on associations with case/control status outside this region.

Associations outside the MHC region

Variants at fifteen loci outside the MHC region showed genome-wide significance (figs. 1, S4, S6; tables 3, S4). Ten of these loci have already been established in previous large MS GWAS (*3*, *4*, *6*). One more locus, *DLEU1*, was only recently confirmed to be associated with MS in a candidate gene study (*21*). The remaining four signals are thus novel candidates for MS susceptibility loci. The lead variants at all fifteen non-MHC loci showed *p*-values $< 5\times10^{-6}$ in DE1 and lower *p*-values $< 5\times10^{-8}$ in the pooled analysis of DE1 and DE2 and have thus replicated in DE2. We could not detect any significant interaction among the 15 top non-MHC variants or between them and SNP rs3104373 within the MHC region.

For validation of our findings, we compared our results to the largest study on MS genetic susceptibility published to date (*4*) (fig. 2). Of the 108 non-MHC variants showing genome-wide

significant or suggestive associations with MS in the published study, 104 variants were present in our data and could be analyzed. All of them showed the same direction of effect (*p*-value of binomial sign test 5×10^{-32}, CI 0.97-1.00), 84 with nominal ($p < 0.05$) and ten with genome-wide significance ($p < 5\times10^{-8}$). 58 of the variants had lower and 35 higher ORs in our data than in the published data set (*4*). It was expected to observe more signals with lower ORs than previously reported due to regression towards the mean.

Next, we examined the four novel loci as well as *DLEU1*, not found at genome-wide significance in a GWAS before, in more detail. We investigated whether the five lead variants at these loci are significantly associated with MS in our German cohort only or whether they replicate in Sardinians, a genetically distinct population with low genetic heterogeneity. This independent Sardinian cohort consisted of 2,903 cases (69.2 % female, 1.2 % PPMS) and 3,323 controls (control/case ratio 1.15) (*22-24*). Two of the variants (rs2812197 and rs4925166) replicated with *p*-values < 0.01 in the Sardinian data set, two more (rs34286592 and rs2836425) showed the same direction of effect but did not reach nominal significance (tables 3, S4; fig. S5).

SHMT1 as a novel MS susceptibility gene

The association of rs4925166 constituted the strongest signal among the novel variants. It showed an OR of 0.85 (CI 0.81-0.90) and a *p*-value of 2.7×10^{-9} in the pooled analysis of German data sets (table 3). This variant replicated in the Sardinian cohort with a joint *p*-value of 7.4×10^{-12} (fig. S4). SNP rs4925166 is located on chromosome 17 in an intron of the gene *TOP3A*, coding for the DNA Topoisomerase III Alpha. However, strongly associated SNPs in this genomic region spread over several neighboring genes (fig. 3A). We therefore conducted an expression quantitative trait locus (eQTL) analysis using a subset of 242 patients from data set DE1 in order to functionally link variants to nearby genes. We examined transcripts within a *cis* window of 1M base pairs up- and downstream of the lead variant for an association of blood gene expression levels with allele configuration (table S5). The variant rs4925166 and proxy SNPs (r^2

> 0.7) were found to be part of a strong eQTL with the gene *SHMT1* in DE1 samples (false discovery rate (FDR) 2.99×10^{-10}, table 4, fig. S8). This eQTL was replicated in two independent control data sets (Max Planck Institute of Psychiatry data (MPIP) (*25*) and Grady Trauma Project (GTP) (*26-28*)) as well as in the publicly available GTEx eQTL database (*29*) (table 4).

To investigate how rs4925166 influences the expression of *SHMT1*, we conducted an association analysis of the SNP with DNA methylation levels in blood. DNA methylation is an important epigenetic mechanism for regulation of gene expression. We tested the association between rs4925166 and DNA methylation levels at CpG sites in the two non-MS data sets MPIP and GTP. Methylation levels at 157 CpG sites that mapped to *SHMT1* were examined for an association with genotype. We observed eight significant (FDR < 0.05) methylation QTLs (mQTLs) between rs4925166 and CpGs in *SHMT1* within the MPIP data set. Three of these associations replicated in the GTP data set (table S6).

We wondered whether the CpG site showing the strongest association with rs4925166 (cg26763362) could fully explain the observed association between the SNP and *SHMT1* expression (causal direction: rs4925166 → cg26763362 → *SHMT1* expression) using mediation analysis (tables 5, S7, S8, figs. 3, S8) (*30*). We observed partial mediation of the effect of rs4925166 on *SHMT1* expression by DNA methylation status of CpG site cg26763362. The association pattern indicates that an additional factor influences the relationship between the SNP, the CpG, and gene expression (see supplementary material). Thus, we conclude that the genotype of rs4925166 affects the expression of *SHMT1* in a complex fashion, partially involving rs4925166-dependent DNA methylation.

Additional novel candidate loci associated with MS

Three loci showed genome-wide significance in the pooled analysis of German data sets DE1 and DE2 but not in Sardinians (table 3). The strongest association, rs4364506, was found on chromosome 6 and is located in an intron of the gene coding for the transcriptional regulator

L3MBTL3 (Lethal(3)malignant brain tumor-like protein 3, fig. S6G). SNP rs2836425 on chromosome 21 constituted the second-strongest signal identified in Germans only. This variant maps to an intron of the gene *ERG*, coding for the transcription factor called V-Ets Avian Erythroblastosis Virus E26 Oncogene Homolog (fig. S6P). Thirdly, SNP rs34286592 is located in an intron of the gene *MAZ* on chromosome 16, coding for the transcription factor MYC-Associated Zink Finger Protein (fig. S6N). It maps to binding sites for transcription factors (fig. S7G).

When conditioning for the lead variants at the four newly identified MS-associated loci, no evidence for secondary signals was found. Thus, the lead variants also constitute the most likely causal variants. These variants all map to introns of genes. This makes a functional link between each variant and the gene it is located in probable. In order to further explore the functional connections between SNPs and genes, we conducted an eQTL analysis of the fifteen loci showing genome-wide significant associations. We thereby identified four cis-eQTLs with FDR < 0.05 in MS cases (table S5). In addition to the eQTL of rs4925166 and *SHMT1* already described above, three more significant eQTLs involved variants at two previously known MS susceptibility loci and three transcripts of the genes *MMEL1* and *ANKRD55*.

Fine-mapping of *DLEU1*

Three variants located on chromosome 13 (rs806321, rs9596270, and rs806349), all intronic within the gene for the long non-coding RNA *DLEU1* (*Deleted in Lymphocytic Leukemia 1*), have been described previously as associated with MS in three large studies (*4-6*), yet the variants did not show genome-wide significance in any of them. The association of rs806349 has recently been confirmed in a candidate-driven follow-up analysis of suggestive MS associations (*21*). However, this variant rs806349 reached a *p*-value of only 2.7×10^{-4} in our analysis (table 5). Instead, a different SNP, rs2812197, in weak LD with rs806349 ($r^2 = 0.4$), showed genome-wide significance in the pooled analysis of DE1 and DE2 and also replicated in Sardinians (tables 3, 5;

fig. S6K). The association of previously described rs806349 is completely dependent on the more strongly associated rs2812197 (table 5). Thus, it is unlikely that rs806349 is the causal SNP at this locus. The same is true for rs806321 (*5*), which is not independent of rs2812197 either (table 5).

The *DLEU1* locus contains evidence for a second signal, rs9591325 (table 5, fig. S6L), in poor LD with rs2812197, but in high LD with rs9596270, which was identified by Patsopoulos *et al.* as a suggestive MS-associated variant (*6*). The two signals were partially independent of each other (table 5). Interestingly, rs9591325 is located in a clear functional region with binding sites for many transcription factors, which is not the case for the other four variants (fig. S7B-F). While rs2812197 shows the overall strongest association at *DLEU1*, the functional data indicates that rs9591325 might be either the actual or a second causal variant. Additional studies with larger sample sizes are required to fully answer this question.

H2: Discussion

The present study constitutes the largest GWAS on MS conducted in a single population to date. By pooled analysis of 3,934 cases in data set DE1 and 954 cases in data set DE2, we identified strong associations in the MHC region with a *p*–value of up to 1.3×10^{-234}. In addition, 15 loci outside the MHC region were associated at a genome-wide significant level (fig. 1, table 3). Associations in the MHC region were examined using imputed *HLA* alleles. Step-wise conditional logistic regression identified *DRB1*15:01* and six more associated *HLA* alleles (table 2), in line with results from previous studies (*2*). All genome-wide significant and suggestive non-MHC MS susceptibility variants published by the IMSGC in 2013 (*4*) and present in our data (n = 104) were replicated regarding direction of effects in our samples (*p*-value 5×10^{-32}, fig. 2).

Four of the 15 non-MHC loci have not been found to be associated with MS in previous studies. One more locus, *DLEU1*, did not reach genome-wide significance in previous GWAS but has

recently been confirmed as MS-associated in a candidate SNP study (*21*). The lead variants at *DLEU1* and at the novel locus *SHMT1* replicated in an independent Sardinian cohort containing 2,903 cases (table 3, fig. S5). Variants at the other three novel loci did not reach nominal significance in Sardinians yet two of them showed the same direction of effect. Due to their consistency and replication within the German cohorts, these three associations can nevertheless be considered as plausible. As the Sardinian population is genetically distinct from Germans, future studies are required to replicate these findings in other cohorts.

Previous genetic analyses of MS susceptibility have indicated immune system related processes as relevant for the development of MS (*4*). Functions of known MS susceptibility genes have been mapped to KEGG pathways JAK-STAT signaling, acute myeloid leukemia (AML), and T cell receptor signaling (*7*). Accordingly, MS-associated genes are predominantly expressed in immune cells (*7*, *8*). The five genes examined in detail in our study (*L3MBTL3*, *DLEU1*, *MAZ*, *ERG*, and *SHMT1*) are associated with regulatory mechanisms in immune cells as well.

The gene *L3MBTL3* encodes a Polycomb-group protein that maintains the transcriptionally repressive state of genes (*31*) and is frequently deleted in several forms of acute leukemia, including AML (*32*). Genes associated with AML constitute one of the most significant pathway categories linked to MS susceptibility variants (*7*). The murine ortholog of *L3MBTL3*, *MBT-1*, has been found to regulate maturation of myeloid progenitor cells (*33*). The regulatory, long non-coding RNA *DLEU1* is often deleted in cases of B-cell chronic lymphocytic leukemia and mantle cell leukemia (*34*). This locus regulates expression of *NF-κB* (*35*), a transcription factor implicated in MS pathology (*4*, *36*, *37*). MAZ is an inflammation-responsive transcription factor (*38*) upregulated during chronic myeloid leukemia (*39*). It binds to the promoter of the gene *MYC*, which is associated with MS (*5*). The transcription factor ERG is important for hematopoiesis (*40*), expression of this oncogene is associated with both AML and acute T-cell lymphoblastic leukemia (*41*). ERG regulates the expression of MS-associated *NF-κB* (*42*), as

DLEU1 does. Finally, SHMT1 is a serine hydroxymethyltransferase acting in the folate cycle. It catalyzes the transfer of a carbon unit subsequently used for synthesis of both nucleotides and methionine. SHMT1 is thus an essential component in the metabolism of the substrate S-adenosylmethionine (SAM), the major methyl group donor during both protein and DNA methylation (*43*, *44*). By this effect on regulation of gene expression, one-carbon metabolism plays an important role in oncogenesis. Lack of *SHMT1* function is, among other effects, associated with acute lymphocytic leukemia (*44-46*). Thus, each of the five genes is involved in regulatory processes of the immune system.

While a clearer picture has already emerged regarding the cell types and broad pathways relevant for the etiology of multiple sclerosis (*3*, *7*), little is still known about the mechanisms by which risk genes act. Analysis of the known functions of the five genes examined in this study revealed that four of them regulate transcription, especially of immune-related genes. Moreover, indirect evidence suggests that they could all be linked either directly or indirectly to epigenetic regulatory mechanisms: L3MBTL3 recognizes epigenetic histone lysine methylation (*31*) and ERG interacts with ESET, a histone H3-specific methyltransferase (*47*). The best known regulatory target of the transcription factor MAZ is *MYC* (*48*), a regulator of epigenetic chromatin state that is associated with MS (*5*, *49*). *DLEU1* is strictly regulated by DNA methylation at its promoter region (*35*). Finally, SHMT1 is essential for maintaining methylation homeostasis in the cell by catalyzing an important reaction in the generation of the methyl donor substrate SAM. Accordingly, establishment of SHMT1 as a MS risk factor puts epigenetic regulation by methylation further in the focus of MS susceptibility.

In recent years, several studies have addressed the role of DNA methylation in the etiology and progression of MS. Methylation differences between MS cases and healthy controls have been analyzed in small, cross-sectional studies. Despite negative results in $CD4^+$ cells (*12*, *50*), Bos and colleagues recently observed significant differences in overall DNA methylation levels in

$CD8^+$ T cells (*12*). Another study demonstrated differentially methylated and expressed genes in brain tissue of MS patients compared to controls (*14*). Furthermore, differential methylation of the major risk locus *HLA-DRB1* was observed in MS patients (*51*). Several groups have found either hyper- or hypomethylation of specific genes to be associated with inflammation or demyelination in MS patients (*11*).

In summary, these studies argue in favor of DNA methylation being relevant for the development of MS. By finding novel risk genes with potential roles in epigenetic regulation, our study adds further indication that epigenetic mechanisms might be important for MS susceptibility. Especially a disturbed homeostasis of methyl donors, caused by an altered expression of *SHMT1*, is likely to have an impact on the disease. As epigenetic mechanisms constitute a major route for environmental risk factors to influence expression of disease-associated genes (*11*), regulation of DNA and protein methylation is an interface where genetic and environmental risk factors for MS might intersect. Detailed analyses of DNA methylation patterns and their interaction with MS susceptibility genes in larger cohorts and among different cell populations and tissues are now required to better understand the role of epigenetic mechanisms in MS.

H2: Materials and Methods

Study samples

Two cohorts of cases, referred to as DE1 and DE2, have been analyzed. Both data sets included patients with CIS, bout onset MS, and primary progressive MS. For cohort DE1, 4,503 cases have been recruited across multiple sites in Germany (for details see the supplementary material). For cohort DE2, 1,002 cases have been recruited across multiple sites in Germany (see supplementary material). The latter cohort has been used in a previous publication (*5*). Controls for these cohorts were obtained from several population-based cohorts across Germany, to match

the different geographical regions where cases were recruited: KORA from the South-Eastern German region of Augsburg (*52*, *53*), HNR from central Western Germany (*54*), SHIP from the North-Eastern region West Pomerania (*55*), DOGS from Dortmund in central Western Germany (*56*), FoCus (*57*) and popgen (*58*) from Kiel in Northern Germany. In addition, controls from two studies on depression conducted in South-Eastern Germany were included (*59, 60*). For a more detailed description of control cohorts see the supplementary material. All responsible ethics committees have provided positive votes for the individual studies. All study participants gave written informed consent. In case of minors, parental informed consent was obtained.

Genotyping and quality control

Samples of cohort DE1 have been genotyped using the Illumina HumanOmniExpress-24-V1-0 or -V1-1 BeadChips. Samples of cohort DE2 have been genotyped using the Illumina Human 660-Quad platform. For both cohorts, identical, stringent QC was conducted on samples and variants. QC steps on samples included removal of individuals with genotyping rate < 2 %, cryptic relatives (relatedness ≥ 1/16), and genetic population outliers. QC steps on variants included removal of variants with call rate < 2 % and MAF < 1%. For a full description of QC, see the supplementary material. Each set of cohorts was combined with controls genotyped on similar arrays, producing case/control data sets DE1 and DE2. QC was repeated on the merged data sets, leading to final figures of 3,934 cases and 8,455 controls for DE1 (table S1), as well as 954 cases and 1,940 controls for DE2 (table S2).

Imputation

Pre-phasing (haplotype estimation) of genotype data was conducted using SHAPEIT2, followed by imputation using IMPUTE2 in 5 Mbp chunks (*16-18*). The 1000 genomes phase I June 2014 release was used as a reference panel. Imputed variants were filtered for MAF ≥ 1 %, INFO metric ≥ 0.8 and HWE p-value ≥ 10^{-6}. For additional details see the supplementary material.

HLA alleles were imputed from genotyping data separately for DE1 and DE2 using HIBAG v1.6.0 (*20*). Alleles with a posterior probability >0.5 were converted to hard calls. Results were validated using *HLA* typing of 442 patients from DE1 (see supplementary material).

Statistical analyses of genotype data

GWAS were conducted on data sets DE1 and DE2 using PLINK2 v1.90b3s (*61*). Sex and the first eight MDS components were used as covariates in logistic regression. Data sets were combined using a fixed-effects model in METASOFT (*62*). For maximum precision, logistic regression and meta analysis of lead SNPs were repeated in R v3.2.3 using package *meta* v4.3.2. All follow-up analyses (e.g., conditional and interaction analyses) were conducted in R. Locus-specific Manhattan plots were generated using LocusZoom with EUR samples of the 1000 genomes March 2012 reference panel on the hg19 build (*63*). For analysis of *HLA* alleles, step-wise logistic regression was conducted in R as described previously (*1*, *2*, *5*).

Gene expression and methylation data

For a subset of 242, mostly treatment-naïve patients from data set DE1 (73 male, 169 female) whole blood RNA was collected using Tempus Blood RNA Tubes (Applied Biosystems, Foster City, CA). RNA was hybridized to Illumina HT-12 v4 expression BeadChips (Illumina, San Diego, CA) and further processed as described in the supplementary material. In summary, QC was conducted in R 3.2.1 using the packages *beadarray* and *lumi* (*64*, *65*). Probes were transformed and normalized through variance stabilization and normalization (VSN) (*66*). Probes which showed a detection p-value < 0.05 in more than 10 % of the samples, which could not be mapped to a known transcript, or which were identified as cross-hybridizing by the Re-Annotator pipeline (*67*) were removed. This left 20,302 transcripts from 242 samples. Technical batch effects were identified by inspecting the association of the first two principal components of expression levels with amplification round, amplification plate, amplification plate column and row, as well as with expression chip. The data were then adjusted using ComBat (*68*).

Gene expression and methylation data of the two control cohorts MPIP (Max Planck Institute of Psychiatry) and GTP (Grady Trauma Project) have been published and described previously and are summarized in the supplementary material (*25*-*28*).

Statistical analysis of gene expression and methylation data

For each of the 15 genome-wide significant loci, all 429 transcripts beginning or ending within 1 Mbp up- or downstream of a lead variant were determined. Associations between genotype and expression levels were examined in data set DE1 by linear regression, using sex, age, and 3 MDS components as covariates. To account for multiple testing, *p*-values were first corrected for the number of transcripts per *cis* window, followed by calculation of the FDR for the total number of variants tested. Replication of eQTLs with an FDR < 0.05 in data set DE1 was conducted in control cohorts MPIP and GTP. For MPIP, the covariates sex, age, BMI, disease status, and 3 MDS components were used in linear regression. For GTP, covariates were sex, age, and 4 MDS components. eQTLs were also looked up in the GTEx database (*29*). Here, only associations in whole blood were considered.

For analysis of the association of rs4925166 with DNA methylation at *SHMT1*, 210 CpG probes were identified in data set MPIP that mapped to *SHMT1*. After removing the quartile of probes showing the lowest variation in methylation status, 157 CpGs remained. Association of DNA methylation with imputed genotype was assessed by linear regression, using sex, age, BMI, disease status, 3 MDS components, and estimated cell counts as covariates. The eight CpG probes showing an FDR < 0.05 were replicated in data set GTP, using sex, age, 4 MDS components, and estimated cell counts as covariates. Mediation analysis was conducted as outlined in the supplementary material, including nonparametric bootstrap for estimation of confidence intervals and *p*-values (*30*).

Replication of the results in a Sardinian cohort

The replication case group consists of the 2,903 unrelated Sardinian MS patients that were diagnosed and selected using the McDonald criteria (*22-24*). Only 35 of these patients were diagnosed with PPMS (1.2 %). 2,010 (69.2 %) cases were female, 893 (30.8 %) male, the average age at onset was 32 years. The matching control group of healthy individuals is composed of 2,880 unrelated adult volunteer blood donors from the same locations where the cases have been collected, as well as 443 Affected Family BAsed pseudo-Controls (AFBAC) derived from 242 MS and 201 type 1 diabetes family trios (*23*). AFBAC allele and haplotype frequencies were constructed using the two alleles in each trio that are not transmitted from the parents to the affected child. These familial pseudo-controls are matched to the cases for ethnic origin and are thus robust to population stratification.

All individuals were genotyped using the Illumina ImmunoChip array. In addition, 2,040 (962 cases and 1,078 controls) were genotyped with the Illumina HumanOmniExpress array and 3,917 (2,111 cases and 1,806 controls) with the Affymetrix 6.0 array. 174 individuals (170 case and 4 controls) were genotyped using both HumanOmniExpress and Affymetrix 6.0 (*22*). After quality control, we used 883,557 SNPs as baseline for imputation (*17*) of 20.1 million untyped SNPs using a Sardinian-specific reference panel including 3,514 Sardinian individuals sequenced to an average coverage of 4.16-fold (*69*).

H2: Supplementary Materials

Results

Fig. S3: Age at onset.

Fig. S4: Forest plots of all top genome-wide significant variants.

Fig. S5: Forest plots of novel variants replicated in a Sardinian cohort.

Fig. S6: Locus-specific Manhattan plots.

Fig. S7: Transcription factor binding sites.

Fig. S8: eQTL and mQTL analysis for rs4925166.

Table S3: Genomic inflation.

Table S4: Genome-wide significant loci.

Table S5: eQTLs with FDR < 0.05 in data set DE1.

Table S6: Replicated mQTLs of rs4925166 and CpG sites in *SHMT1*.

Table S7: Mediation analysis.

Table S8: Causal mediation analysis.

Materials and Methods

Fig. S1: Substructure analysis results in DE1.

Fig. S2: Substructure analysis results in DE2.

Table S1: Quality control of data set DE1.

Table S2: Quality control of data set DE2.

H2: References and Notes

1. N. A. Patsopoulos *et al.*, Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet*. **9**, e1003926 (2013).

2. International Multiple Sclerosis Genetics Consortium, International IBD Genetics Consortium (IIBDGC), Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nature Genetics*. **47**, 1107–1113 (2015).

3. S. Sawcer, R. J. M. Franklin, M. Ban, Multiple sclerosis genetics. *Lancet Neurol*. **13**, 700–709 (2014).

4. International Multiple Sclerosis Genetics Consortium (IMSGC), Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics*. **45**, 1353–1360 (2013).

5. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. **476**, 214–219 (2011).

6. N. A. Patsopoulos *et al.*, Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol.* **70**, 897–912 (2011).

7. International Multiple Sclerosis Genetics Consortium, Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. *Am. J. Hum. Genet.* **92**, 854–865 (2013).

8. K. K.-H. Farh *et al.*, Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. **518**, 337–343 (2015).

9. C. H. Hawkes, A. J. Macgregor, Twin studies and the heritability of MS: a conclusion. *Mult. Scler.* **15**, 661–667 (2009).

10. A. K. Hedström, T. Olsson, L. Alfredsson, The Role of Environment and Lifestyle in Determining the Risk of Multiple Sclerosis. *Curr Top Behav Neurosci*. **26**, 87–104 (2015).

11. Y. Zhou *et al.*, The potential role of epigenetic modifications in the heritability of multiple sclerosis. *Mult. Scler.* **20**, 135–140 (2014).

12. S. D. Bos *et al.*, Genome-wide DNA methylation profiles indicate CD8+ T cell hypermethylation in multiple sclerosis. *PLoS ONE*. **10**, e0117403 (2015).

13. J. L. Huynh, P. Casaccia, Epigenetic mechanisms in multiple sclerosis: implications for pathogenesis and treatment. *Lancet Neurol*. **12**, 195–206 (2013).

14. J. L. Huynh *et al.*, Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat Neurosci*. **17**, 121–130 (2014).

15. M. Steffens *et al.*, SNP-based analysis of genetic substructure in the German population. *Hum. Hered.* **62**, 20–29 (2006).

16. B. N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. **5**, e1000529 (2009).

17. B. Howie, C. Fuchsberger, M. Stephens, J. Marchini, G. R. Abecasis, Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*. **44**, 955–959 (2012).

18. O. Delaneau, J.-F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods*. **10**, 5–6 (2013).

19. M. L. Freedman *et al.*, Assessing the impact of population stratification on genetic association studies. *Nature Genetics*. **36**, 388–393 (2004).

20. X. Zheng *et al.*, HIBAG--HLA genotype imputation with attribute bagging. *The Pharmacogenomics Journal*. **14**, 192–200 (2014).

21. C. M. Lill *et al.*, Genome-wide significant association with seven novel multiple sclerosis risk loci. *J. Med. Genet.* **52**, 848–855 (2015).

22. S. Sanna *et al.*, Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nature Genetics*. **42**, 495–497 (2010).

23. M. Zoledziewska *et al.*, Variation within the CLEC16A gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in Sardinia. *Genes and Immunity*. **10**, 15–17 (2009).

24. N. Barizzone *et al.*, The burden of multiple sclerosis variants in continental Italians and Sardinians. *Mult. Scler.* **21**, 1385–1395 (2015).

25. J. Arloth *et al.*, Genetic Differences in the Immediate Transcriptome Response to Stress Predict Risk-Related Brain Function and Psychiatric Disorders. *Neuron*. **86**, 1189–1202 (2015).

26. E. B. Binder *et al.*, Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. *JAMA*. **299**, 1291–1305 (2008).

27. D. Mehta *et al.*, Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc Natl Acad Sci USA*. **110**, 8302–8307 (2013).

28. A. S. Zannas *et al.*, Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol*. **16**, 266 (2015).

29. M. Melé *et al.*, Human genomics. The human transcriptome across tissues and individuals. *Science*. **348**, 660–665

(2015).

30. D. Tingley, T. Yamamoto, K. Hirose, L. Keele, K. Imai, mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*. **59**, 1–38 (2014).

31. N. Nady *et al.*, Histone recognition by human malignant brain tumor domains. *J. Mol. Biol.* **423**, 702–718 (2012).

32. M. Merup *et al.*, 6q deletions in acute lymphoblastic leukemia and non-Hodgkin's lymphomas. *Blood*. **91**, 3397–3400 (1998).

33. S. Arai, T. Miyazaki, Impaired maturation of myeloid progenitors in mice lacking novel Polycomb group protein MBT-1. *EMBO J*. **24**, 1863–1873 (2005).

34. S. Stilgenbauer *et al.*, Expressed sequences as candidates for a novel tumor suppressor gene at band 13q14 in B-cell chronic lymphocytic leukemia and mantle cell lymphoma. *Oncogene*. **16**, 1891–1897 (1998).

35. A. Garding *et al.*, Epigenetic upregulation of lncRNAs at 13q14.3 in leukemia is linked to the In Cis downregulation of a gene cluster that targets NF-kB. *PLoS Genet*. **9**, e1003373 (2013).

36. J. Yan, J. M. Greer, NF-kappa B, a potential therapeutic target for the treatment of multiple sclerosis. *CNS Neurol Disord Drug Targets*. **7**, 536–557 (2008).

37. W. J. Housley *et al.*, Genetic variants associated with autoimmunity drive NFκB signaling and responses to inflammatory stimuli. *Sci Transl Med*. **7**, 291ra93 (2015).

38. A. Ray *et al.*, SAF-3, a novel splice variant of the SAF-1/MAZ/Pur-1 family, is expressed during inflammation. *FEBS J*. **276**, 4276–4286 (2009).

39. L. Dahéron *et al.*, Identification of several genes differentially expressed during progression of chronic myelogenous leukemia. *Leukemia*. **12**, 326–332 (1998).

40. S. J. Loughran *et al.*, The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat. Immunol.* **9**, 810–819 (2008).

41. C. D. Baldus *et al.*, High expression of the ETS transcription factor ERG predicts adverse outcome in acute T-lymphoblastic leukemia in adults. *J. Clin. Oncol.* **24**, 4714–4720 (2006).

42. B. Hoesel, J. A. Schmid, The complexity of NF-κB signaling in inflammation and cancer. *Mol. Cancer*. **12**, 86 (2013).

43. K. S. Crider, T. P. Yang, R. J. Berry, L. B. Bailey, Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. *Adv Nutr*. **3**, 21–38 (2012).

44. J. W. Locasale, Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat. Rev. Cancer*. **13**, 572–583 (2013).

45. R. de Jonge *et al.*, Polymorphisms in folate-related genes and risk of pediatric acute lymphoblastic leukemia. *Blood*. **113**, 2284–2289 (2009).

46. C. F. Skibola *et al.*, Polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and risk of adult acute lymphocytic leukemia. *Blood*. **99**, 3786–3791 (2002).

47. L. Yang *et al.*, Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. *Oncogene*. **21**, 148–152 (2002).

48. S. A. Bossone, C. Asselin, A. J. Patel, K. B. Marcu, MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc Natl Acad Sci USA*. **89**, 7452–7456 (1992).

49. N. V. Varlakhanova, P. S. Knoepfler, Acting locally and globally: Myc's ever-expanding roles on chromatin. *Cancer Res.* **69**, 7487–7490 (2009).

50. S. E. Baranzini *et al.*, Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature*. **464**, 1351–1356 (2010).

51. M. Graves *et al.*, Methylation differences at the HLA-DRB1 locus in CD4+ T-Cells are associated with multiple sclerosis. *Mult. Scler.* **20**, 1033–1041 (2013).

52. R. Holle, M. Happich, H. Löwel, H. E. Wichmann, MONICA/KORA Study Group, KORA--a research platform for population based health research. *Gesundheitswesen*. **67 Suppl 1**, S19–25 (2005).

53. H. E. Wichmann, C. Gieger, T. Illig, MONICA/KORA Study Group, KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*. **67 Suppl 1**, S26–30 (2005).

54. A. Schmermund *et al.*, Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am. Heart J.* **144**, 212–218 (2002).

55. H. Völzke *et al.*, Cohort profile: the study of health in Pomerania. *Int J Epidemiol*. **40**, 294–307 (2011).

56. K. Berger, DOGS. *Bundesgesundheitsbl.* **55**, 816–821 (2012).

57. N. Müller *et al.*, IL-6 blockade by monoclonal antibodies inhibits apolipoprotein (a) expression and lipoprotein (a) synthesis in humans. *J. Lipid Res.* **56**, 1034–1042 (2015).

58. M. Krawczak *et al.*, PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet*. **9**, 55–61 (2006).

59. P. Muglia *et al.*, Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol. Psychiatry*. **15**, 589–601 (2010).

60. M. A. Kohli *et al.*, The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron*. **70**, 252–265 (2011).

61. C. C. Chang *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. **4**, 7 (2015).

62. B. Han, E. Eskin, Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).

63. R. J. Pruim *et al.*, LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. **26**, 2336–2337 (2010).

64. M. J. Dunning, M. L. Smith, M. E. Ritchie, S. Tavaré, beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics*. **23**, 2183–2184 (2007).

65. P. Du, W. A. Kibbe, S. M. Lin, lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. **24**, 1547–1548 (2008).

66. W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*. **18 Suppl 1**, S96–104 (2002).

67. J. Arloth, D. M. Bader, S. Röh, A. Altmann, Re-Annotator: Annotation Pipeline for Microarray Probe Sequences. *PLoS ONE*. **10**, e0139516 (2015).

68. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes

methods. *Biostatistics*. **8**, 118–127 (2007).

69. C. Sidore *et al.*, Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nature Genetics*. **47**, 1272–1281 (2015).

70. M. B. Gerstein *et al.*, Architecture of the human regulatory network derived from ENCODE data. *Nature*. **489**, 91–100 (2012).

71. J. Wang *et al.*, Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res*. **41**, D171–6 (2013).

72. J. Wang *et al.*, Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res*. **22**, 1798–1812 (2012).

73. M. J. Aryee *et al.*, Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. **30**, 1363–1369 (2014).

74. J.-P. Fortin *et al.*, Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol*. **15**, 503 (2014).

75. E. A. Houseman *et al.*, DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. **13**, 86 (2012).

H2: Acknowledgments

H3: General:

We thank Verena Grummel, Joachim Hornung, and Nadine Miksch for technical assistance as well as Nadine Provençal and Stephan Ripke for scientific discussions. This study makes use of data generated by the Wellcome Trust Case-Control Consortium (WTCCC). A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk.

H3: Funding:

This work was supported by the German Ministry for Education and Research (BMBF) as part of the “German Competence Network Multiple Sclerosis” (KKNMS) (grant numbers 01GI0916, 01GI0917) and the Munich Cluster for Systems Neurology (SyNergy). Dorothea Buck, Bernhard Hemmer, Frauke Zipp and Heinz Wiendl were supported by the German Research foundation (DFG) (grant no. CRC128). Funding for the WTCCC project was provided by the Wellcome Trust under award 076113 and 085475. The collection of sociodemographic and clinical data in the Dortmund Health Study was supported by the German Migraine & Headache Society

(DMKG) and by unrestricted grants of equal share from Almirall, Astra Zeneca, Berlin Chemie, Boehringer, Boots Health Care, Glaxo-Smith-Kline, Janssen Cilag, McNeil Pharma, MSD Sharp & Dohme, and Pfizer to the University of Münster. Blood collection in the Dortmund Health Study was done through funds from the Institute of Epidemiology and Social Medicine University of Münster; genotyping was supported by the BMBF (grant no. 01ER0816). The FoCus study was supported by the BMBF (grant no. 0315540A). The Heinz Nixdorf Recall Study was supported by the Heinz Nixdorf Foundation Germany, the BMBF, and the DFG (ER 155/6-1, ER 155/6-2). The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the BMBF and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The popgen 2.0 network is supported by a grant from the BMBF (grant no. 01EY1103). SHIP is part of the Community Medicine Research Network of the University Medicine Greifswald (www.community-medicine.de), which was initiated and funded by the BMBF and the State of Mecklenburg-Pomerania; genome-wide data have been supported by the BMBF (grant no. 03ZIK012).

H3: Author contributions:

Conception and design: TFMA, DB, BH, BMM.

Recruitment of German cases: DB, GA, ABa, LB, ABe, AC, CGa, RG, CGr, JH, MH, CID, MK, TKü, VLi, RAL, VLo, FL, SGM, MM, SN, FP, MPü, TR, AS, MSt, JPS, KHS, BT, FTB, HT, CW, FW, HW, BW, UKZ, UZ, FZ, BH.

Acquisition of genotype or expression data of German cases (DE1): PW, MRB, PL, TB, FW.

German control cohorts: KB, LB, AF, CG, SH, GH, MI, KHJ, SK, TKa, ML, WL, CML, SL, TM, SM, MMüN, MMNö, AP, RR, KS, US, HV, MW, JW.

Provided replication data (Sardinian cohort): AM, MPi, EP, CS, IZ, FC, MZ.

Analysis and interpretation of data: TFMA, DB, JA, MOS, TD, AZ, DC, TCR, EBB, BH, BMM. Drafting or revising the manuscript: TFMA, DB, FC, MZ, BH, BMM.

H3: Competing interests:

A. Bayas received honoraria for consultancy and/or as speaker from Merck Serono, Biogen, Bayer Vital, Novartis, TEVA, Roche and Sanofi/Genzyme; for trial activities from Biogen, Merck Serono and Novartis; he received grants for congress trips and participation from Biogen, Novartis, Sanofi/Genzyme, and Merck Serono. A. Berthele received travel grants, research grants, and speaker honoraria from Bayer Healthcare, Biogen, Merck Serono, Novartis, Teva, and Sanofi. D. Buck received compensation for activities with Bayer HealthCare, Biogen, MerckSerono, and Novartis; she is supported by the Abirisk Consortium. A. Chan received compensation for activities with Almirall Hermal GmbH, Bayer Schering, Biogen Idec, Merck Serono, Novartis and Teva Neuroscience, research support from Bayer Schering, Biogen Idec, Merck Serono and Novartis, and research grants from the BMBF (KKNMS, CONTROL MS, 01GI0914). R. Gold received compensation for activities with Bayer Healthcare, Biogen Idec and Teva Neuroscience, and for an editorial capacity from Therapeutic Advances in Neurological Disorders, patent payments from Biogen Idec, and research support from Bayer Healthcare, Biogen Idec, Merck Serono, Teva Neuroscience, Novartis and from the BMBF (KKNMS, CONTROL MS, 01GI0914). M. Hecker received speaker honoraria and travel expenses from Bayer HealthCare, Biogen Idec, Novartis, and Teva. B. Hemmer served on scientific advisory boards for Roche, Novartis, Bayer Schering, Merck Serono, Biogen Idec, GSK, Chugai Pharmaceuticals, Genentech and Genzyme Corporation; he serves on the international advisory board of Archives of Neurology and Experimental Neurology; he received speaker honoraria from Bayer Schering, Novartis, Biogen Idec, Merck Serono, Roche, and Teva Pharmaceutical Industries; he received research support from Biogen Idec, Bayer Schering, Merck Serono, Five prime, Metanomics, Chugai Pharmaceuticals, and Novartis. He has been

filed a patent for the detection of antibodies and T cells against KIR4.1 in a subpopulation of MS patients and genetic determinants of neutralizing antibodies to interferon-beta. M. Knop received travel grants for attention of scientific meetings from Genzyme and grant support from Merck Serono. F. Luessi received travel grants from Teva Pharma and Merck Serono. S.G. Meuth received honoraria for lecturing and travel expenses for attending meetings and has received financial research support from Bayer, Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Merck Sharp & Dohme, Novartis, Novo Nordisk, Sanofi-Aventis, and Teva. M. Mühlau received research support from BMBF, DFG, Hertie Foundation, Merck Serono and Novartis, and travel expenses for attending meetings from Bayer and Merck Serono; he received honoraria for lecturing from Merck Serono, and investigator fees for a Phase III clinical study from Biogen Idec. S. Nischwitz received travel grants for attention of scientific meetings from Merck Serono, Biogen Idec and TEVA, and grant support from Bayer Schering and Novartis. T. Ruck received travel expenses and financial research support from Genzyme and honoraria for lecturing from Teva and Genzyme. A. Salmen received personal compensation for activities with Novartis, Sanofi, and Almirall Hermal GmbH. U. Schminke's research activities were funded by the National Institute of Neurological Disorders and Stroke (sub award #A08580 M10A10647), the BMBF (grant #03IS2061A), the German Federal State of Mecklenburg-West Pomerania, and Siemens Healthcare. He received travel expenses and/or honoraria for lectures or educational activities not funded by industry. J.P. Stellmann and K.H. Stürner received research grants and speaker honoraria from Bayer Healthcare, Biogen, Merck Serono, Novartis, and Sanofi Aventis. M. Stangel received honoraria for scientific lectures or consultancy from Bayer Healthcare, Biogen Idec, Baxter, CSL Behring, Grifols, Merck-Serono, Novartis, Sanofi-Aventis, and Teva. His institution received research support from Bayer Healthcare, Biogen Idec, Merck-Serono, Novartis, and Teva. His lab has grant support from the DFG, the Ministry of Science and Culture of Lower Saxony (N-RENNT), the BMBF, and the Röver foundation. B. Tackenberg received

consultancy and speaker honoraria and/or research grants from Bayer Healthcare, Biogen, CSL Behring, Genzyme, Grifols, Merck-Serono, Novartis, Octapharma, Roche, Sanofi-Aventis, and Teva. H. Tumani received honoraria for speaking/consultation and travel grants from Bayer Healthcare, Biogen Idec, Merck Serono, Genzyme, Novartis Pharma, Siemens Health Products, and Teva Pharma and research grants from Biogen Idec, Merck Serono, Novartis Pharma, Siemens Health Products, Teva Pharma, and the BMBF. C. Warnke received honoraria for participation to advisory boards and/or research funding from Novartis, Bayer, Biogen and TEVA. F. Weber received honoraria from Genzyme and Novartis for serving on a scientific advisory board and a travel grant for the attention of a scientific meeting from Merck-Serono, Novartis, and Biogen. He received grant support from Merck-Serono, Novartis, and the BMBF (projects Biobanking and Omics in CONTROL MS, KKNMS). H. Wiendl received compensation for serving on Scientific Advisory Boards/Steering Committees for Bayer Healthcare, Biogen Idec, Genzyme, Merck Serono, Novartis, and Sanofi Aventis. He received speaker honoraria and travel support from Bayer Vital GmbH, Bayer Schering AG, Biogen Idec, CSL Behring, EMD Serono, Fresenius Medical Care, Genzyme, Merck Serono, Omniamed, Novartis, and Sanofi Aventis, and compensation as a consultant from Biogen Idec, Merck Serono, Novartis, and Sanofi Aventis. He received research support from Bayer Vital, Biogen Idec, Genzyme Merck Serono, Novartis, Sanofi Aventis Germany, Sanofi US as well as grants and research support from Bayer Healthcare, Biogen Idec, BMBF, DFG, Else Kröner Fresenius Foundation, Fresenius Foundation, Hertie Foundation, Merck Serono, Novartis, NRW Ministry of Education and Research, Interdisciplinary Center for Clinical Studies (IZKF) Münster, RE Children's Foundation, Sanofi Aventis/Genzyme, and TEVA Pharma. B. Wildemann received honoraria for speaking/consultation and travel grants from Bayer Healthcare, Biogen Idec, Merck Serono, Genzyme, a Sanofi Company, Novartis Pharmaceuticals, and Teva Pharma GmbH and research grants from Biogen Idec, Biotest, Merck Serono, Novartis Pharmaceuticals,

Teva Pharma GmbH, the BMBF, and the Dietmar Hopp Foundation. U. Ziemann received honoraria for speaking/consultation from Bayer Health Care, Biogen Idec, Bristol Myers Squibb, CorTec, Medtronic GmbH, Servier, and research funding from Biogen Idec.

H3: Data and materials availability:

All data needed to evaluate the conclusions in the paper are present in the paper and the supplementary materials or are available from authors upon request.

H2: Figures and Tables

Fig. 1. Genome-wide representation of MS associations in the pooled analysis of German data sets

Manhattan plot showing strength of evidence for association (*p*-value). Each variant is shown as a dot, with alternating shades of blue according to chromosome. Green dots represent established MS-associated variants and their proxies, as listed by Sawcer *et al.* (*3*) (except for rs2812197, which was not covered by that review). Top variants at the 15 non-MHC loci associated at the genome-wide significance threshold in our study are shown as diamonds. Novel variants showing genome-wide significance are plotted as red diamonds, their names are shown in bold font. Variants in high linkage disequilibrium ($r^2 \geq 0.7$) with these novel variants are shown as red dots. Variants replicating in the Sardinian cohort are shown in red font. MA = minor allele, OR = odds ratio (relative to the MA). Gene names for known loci are indicated as listed by Sawcer *et al.* (*3*). The plot is truncated at $-\log_{10}(p) = 16$ for better visibility, all truncated variants map to the MHC region. The lowest *p*-value (rs3104373, *) was 1.3×10^{-234}.

Fig. 2. Comparison of results from the pooled analysis of Germans to associations found in an IMSGC study.

104 of the 108 variants showing genome-wide significant or suggestive associations with MS in the study published by the IMSGC in 2013 (*4*) were present in the pooled results of DE1 and DE2. All 104 variants showed the same direction of effect (binomial test *p*-value 5×10^{-32}). 58 variants had lower and 35 higher ORs compared to the published data set. *P*-value-based categories labeled with different dots represent exclusive bins that add up to 104. CI = 95 % confidence interval, OR = odds ratio.

Fig. 3. Fine-mapping analysis results of locus rs4925166.

(A): Regional plot for the rs4925166/*SHMT1* locus. Color of dots indicates LD with the lead variant (rs4925166, shown in pink). Grey dots represent signals with missing r^2 values.

(**B):** Mediation analysis results in MPIP/GTP controls. Mediation effect: rs4925166 → CpG cg26763362 → *SHMT1* expression. Direct effect: rs4925166 → *SHMT1* expression. Data has been calculated using the R package *mediation* (*30*), except for total effect (*), which was calculated by linear regression. Results were obtained using 1,000,000 simulations. Effects and *p*-values shown here differ from table 5, as a lower number of samples contained both expression and methylation data than expression data alone. **(C):** Relationship between *cg26763362* methylation, *SHMT1* expression, and rs4925166 genotype in MPIP controls.

	Cohort DE1	Cohort DE2
Number of cases	3934	954
Age [mean (range)]	39 (13-79)	40 (17-82)
Female [n (%)]	2723 (69.2)	695 (72.9)
Male [n (%)]	1211 (30.8)	259 (27.1)
PPMS [n (%)]	105 (2.7)	63 (6.6)

Table 1: Clinical characteristics of German MS cases.

PPMS = Primary progressive MS (as opposed to bout onset MS).

***HLA* allele**	**AF**	**OR (95 % CI)**	***p*-value**	***HLA* alleles in LD (r^2 > 0.9)**
DRB1*15:01	14.8	2.85 (2.66-3.06)	1.03×10^{-191}	DQB1*06:02
A*02:01	28.6	0.68 (0.64-0.73)	3.68×10^{-29}	
B*38:01	2.0	0.36 (0.27-0.49)	2.09×10^{-11}	
DRB1*13:03	1.5	1.96 (1.60-2.40)	6.42×10^{-11}	
DPB1*03:01	10.3	1.33 (1.22-1.46)	4.35×10^{-10}	
DRB1*03:01	12.2	1.29 (1.18-1.40)	1.85×10^{-08}	DQA1*05:01, DQB1*02:01
DRB1*08:01	3.0	1.63 (1.39-1.91)	2.36×10^{-09}	DQA1*04:01, DQB1*04:02

Table 2: Genome-wide significant *HLA* alleles.

Alleles are in order of step-wise logistic regression. For each row, alleles from the rows above have been used as covariates in the model. AF (allele frequency of controls in %) is calculated from a joint set of DE1 and DE2. ORs and *p*-values are from a fixed-effects pooled analysis of DE1 and DE2.

Variant	C	MA	Gene	MAF DE	OR (CI) DE1+DE2	*p*-value DE1+DE2	*p*-value Sardinia	OR (CI) DE+Sard.	*p*-value DE+Sard.
rs10797431	1	T	*MMEL1*	34.1	0.84 (0.80-0.89)	1.81×10^{-10}			
rs6689470	1	A	*EVI5*	14.2	1.24 (1.16-1.33)	3.93×10^{-10}			
rs2300747	1	G	*CD58*	12.4	0.75 (0.69-0.81)	1.74×10^{-12}			
rs7535818	1	G	*RGS1*	19.2	0.76 (0.71-0.82)	1.51×10^{-15}			
rs2681424	3	C	*CD86*	49.7	0.86 (0.82-0.90)	9.51×10^{-10}			
rs6859219	5	A	*ANKRD55*	22.2	0.84 (0.79-0.89)	8.06×10^{-09}			
rs3104373	6	T	*HLA-DRB1*	13.6	2.90 (2.72-3.09)	1.34×10^{-234}			
rs4364506	**6**	**A**	***L3MBTL3***	**26.4**	**0.84 (0.80-0.89)**	$\mathbf{4.06\times10^{-09}}$	0.83	0.89 (0.85-0.93)	1.99×10^{-06}
rs2182410	10	T	*IL2RA*	38.1	0.84 (0.79-0.88)	1.15×10^{-11}			
rs1891621	10	G	Intergenic	46.7	0.87 (0.83-0.91)	2.94×10^{-08}			
rs1800693	12	C	*TNFRSF1A*	42.1	1.17 (1.11-1.23)	1.06×10^{-09}			
rs2812197	13	T	*DLEU1*	38.4	0.86 (0.82-0.91)	9.95×10^{-09}	$\mathbf{6.86\times10^{-03}}$	**0.87 (0.83-0.91)**	$\mathbf{2.83\times10^{-10}}$
rs6498168	16	T	*CLEC16A*	35.5	1.23 (1.17-1.29)	1.98×10^{-15}			
rs34286592	**16**	**T**	***MAZ***	**14.2**	**1.21 (1.13-1.30)**	$\mathbf{4.58\times10^{-08}}$	0.44	1.16 (1.09-1.23)	4.79×10^{-07}
rs4925166	**17**	**T**	***SHMT1***	**34.5**	**0.85 (0.81-0.90)**	$\mathbf{2.69\times10^{-09}}$	$\mathbf{5.63\times10^{-04}}$	**0.86 (0.82-0.90)**	$\mathbf{7.40\times10^{-12}}$
rs2836425	**21**	**T**	***ERG***	**12.7**	**1.22 (1.14-1.31)**	$\mathbf{2.84\times10^{-08}}$	0.35	1.18 (1.11-1.25)	1.54×10^{-07}

Table 3: Genome-wide significant loci outside the MHC region and the top variant within the MHC region.

Bold font in the left half of the table indicates novel loci, bold font in the right half variants that replicated in Sardinians. All *p*-values shown in the table are two-sided. Gene names of known loci are as listed by Sawcer *et al.* (*3*). C = Chromosome, MA = Minor allele. For additional details see table S4.

Expression				
Data set	**Transcript**	**Effect**	***p*-value**	**FDR**
DE1	*SHMT1*	0.36	4.42×10^{-13}	2.99×10^{-10}
MPIP	*SHMT1*	0.19	4.26×10^{-12}	1.28×10^{-11}
GTP	*SHMT1*	0.11	3.12×10^{-04}	1.25×10^{-03}
GTEx	*SHMT1*	0.56	9.2×10^{-28}	NA
Methylation				
Data set	**CpG**	**Effect**	***p*-value**	**FDR**
MPIP	*cg26763362*	-0.03	3.21×10^{-20}	5.04×10^{-18}
GTP	*cg26763362*	-0.03	1.98×10^{-14}	1.58×10^{-13}

Table 4: eQTL and mQTL analysis for rs4925166.

Direction of effect is relative to the minor allele T. Note that the effect sizes cannot be directly compared as normalization methods and covariates partly differ between studies. Additional eQTLs and mQTLs are described in the supplementary material. As only the single eQTL rs4925166/*SHMT1* was examined in GTEx data, no FDR is indicated here.

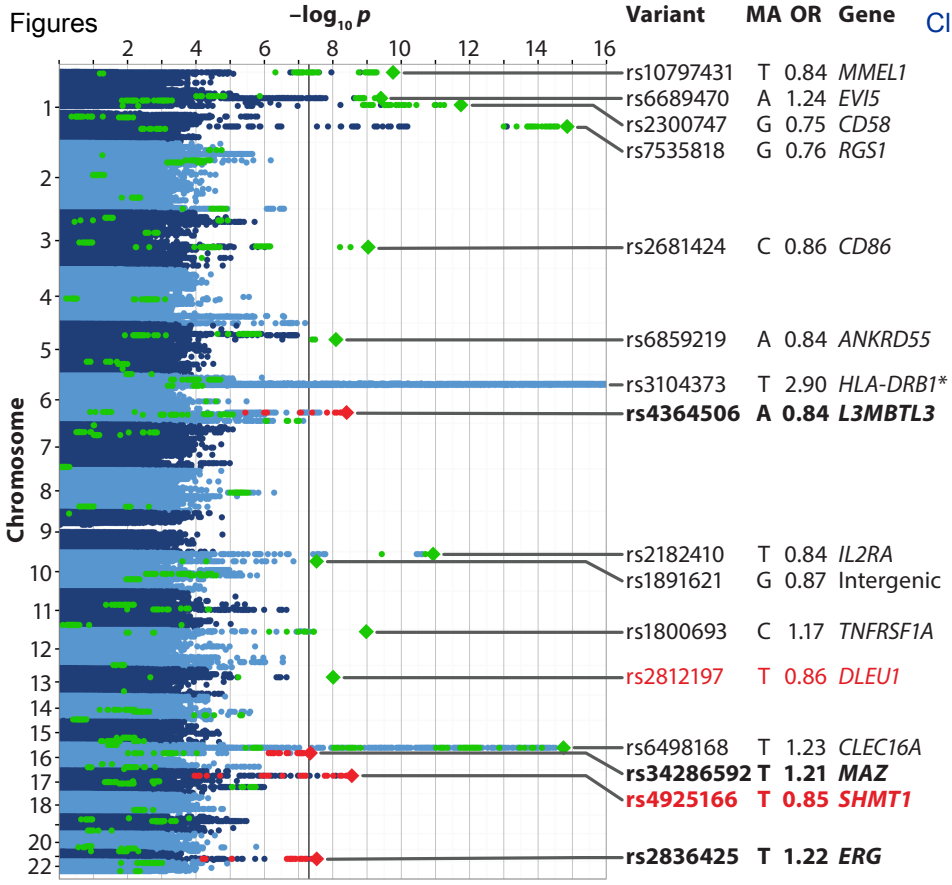
Variant	**MAF**	**OR (CI)**	***p*-value**	***p*-value** (rs281297)	***p*-value** (rs9591325)	r^2	**Ref.**
rs2812197	38.4	0.86 (0.82-0.91)	9.95×10^{-9}		4.79×10^{-5}	1.00	
rs806321	48.5	0.89 (0.85-0.94)	6.36×10^{-6}	0.81	2.02×10^{-3}	0.66	(*5*)
rs806349	46.0	1.10 (1.04-1.15)	2.73×10^{-4}	0.99	0.019	0.41	(*4*, *21*)
rs9591325	8.1	0.78 (0.70-0.85)	2.26×10^{-7}	9.13×10^{-4}		0.14	
rs9596270	8.1	0.78 (0.71-0.86)	4.45×10^{-7}	1.49×10^{-3}	0.27	0.14 (0.99†)	(*6*)

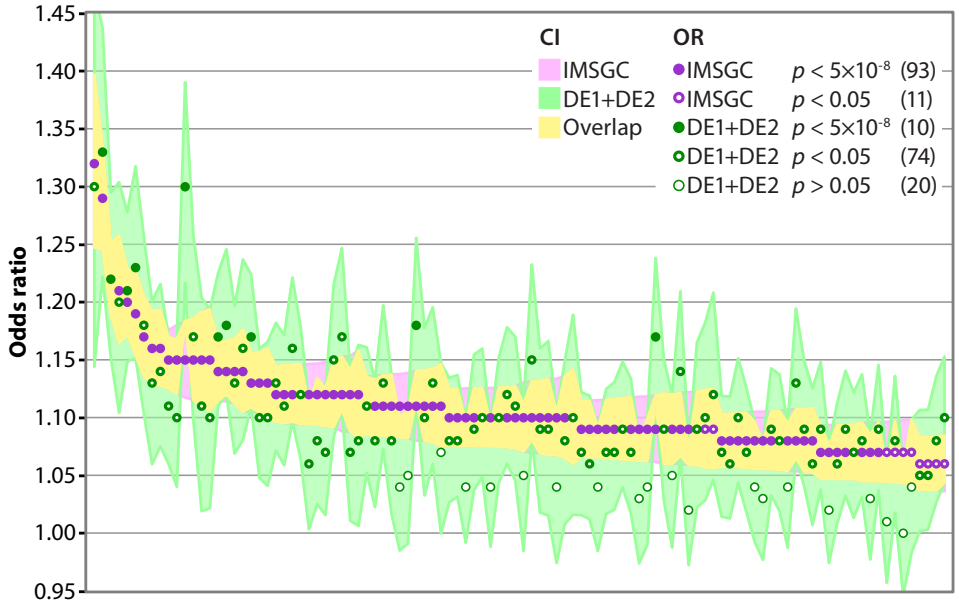
Table 5: Fine-mapping of the *DLEU1* locus.

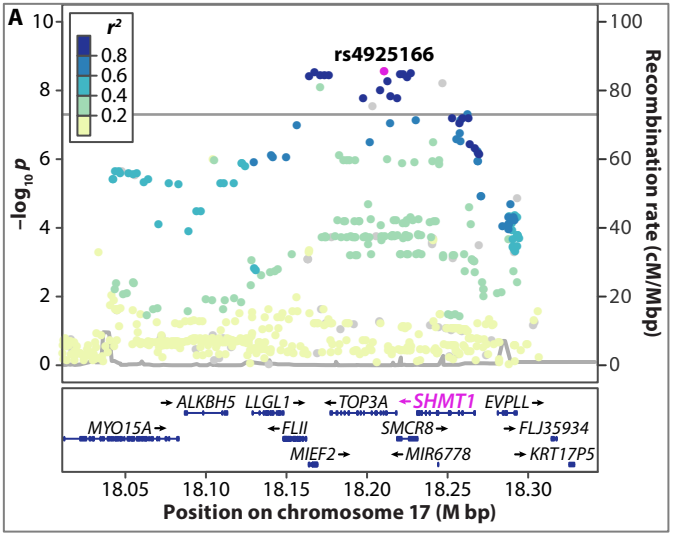
MAF (controls in %) and r^2 (with rs2812197) are calculated from a joint set of DE1 and DE2, ORs and *p*-values from the pooled analysis of DE1 and DE2. Second and third *p*-value columns are from conditional analysis. †: r^2 with rs9591325.

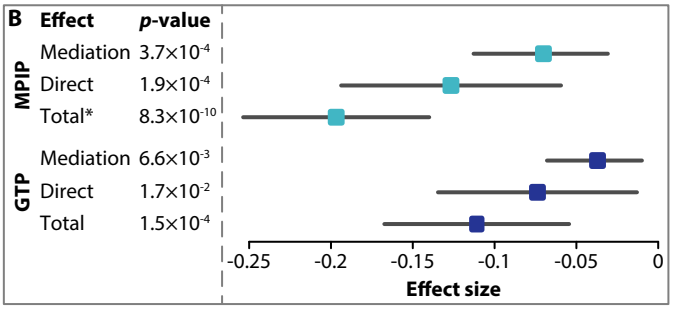
Figures

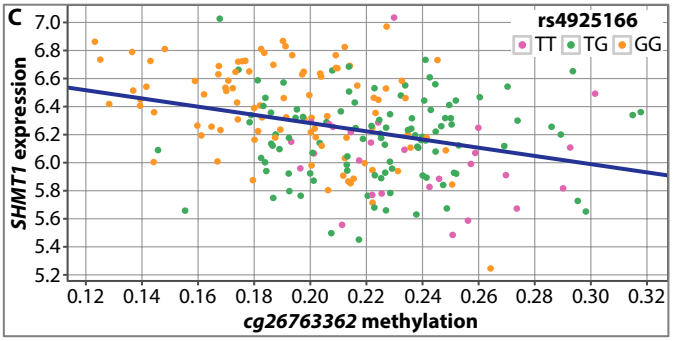
Click here to download Figure AndlauerFigures.pdf




Odds ratio
1.45
1.40
1.35
1.30
1.25
1.20
1.15
1.10
1.05
1.00
0.95
CI
IMSGC
DE1+DE2
Overlap
OR
IMSGC $p < 5\times10^{-8}$ (93)
IMSGC $p < 0.05$ (11)
DE1+DE2 $p < 5\times10^{-8}$ (10)
DE1+DE2 $p < 0.05$ (74)
DE1+DE2 $p > 0.05$ (20)


A
rs4925166
r²
0.8
0.6
0.4
0.2
−log10 p
Recombination rate (cM/Mbp)
ALKBH5
LLGL1
TOP3A
SHMT1
EVPLL
MYO15A
FLII
SMCR8
FLJ35934
MIEF2
MIR6778
KRT17P5
Position on chromosome 17 (M bp)


B
Effect
p-value
MPIP
Mediation 3.7×10-4
Direct 1.9×10-4
Total* 8.3×10-10
GTP
Mediation 6.6×10-3
Direct 1.7×10-2
Total 1.5×10-4
Effect size


C
rs4925166
TT
TG
GG
SHMT1 expression
cg26763362 methylation

Supplementary Tables S4 and S5